

Homework Assignment 5, Due Tuesday December 2, 2008

CSCI 6490/4490 Algorithms for Computational Biology

November 19, 2008

1. Problem 11.4 in Chapter 11 on page 407. Note that you need to build a dynamic programming table to compute log-scores for all prefixes of **GGCT** and to record the option used to compute every cell in the table. You also need to indicate the path corresponding a most likely sequence of states for **GGCT**.
2. Design a *profile* HMM for the profile given in Figure 4.3 on page 96. Draw the profile HMM you design including all the transition and emission probabilities.
3. Continue for question 2. This time, you need to assume pseudo counts for emission of nucleotides in **MATCH** states only. You will need to sequence information in Figure 4.3 instead of the profile. You come up with some pseudo counts and give the rationale. Again draw the profile HMM you design including all the transition and emission probabilities.
4. Continue for questions 2 and 3. This time, you need to consider pseudo counts for insertions and deletions. You come up with some pseudo counts and give the rationale. Again draw the profile HMM you design including all the transition and emission probabilities.